# Multilocus sequence typing of *Enterocytozoon bieneusi*: Lack of geographic segregation and existence of genetically isolated sub-populations ☆

Wei Li [a,b], Vitaliano Cama [c], Frederick O. Akinbo [d], Sandipan Ganguly [e], Nicholas M. Kiulia [f], Xichen Zhang [g], Lihua Xiao [a,*]

[a] *Division of Foodborne, Waterborne and Environmental Diseases, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA*
[b] *College of Veterinary Medicine, Northeast Agricultural University, Harbin 150030, China*
[c] *Division of Parasitic Diseases and Malaria, Center for Global Health, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA*
[d] *Department of Medical Laboratory Science, University of Benin, Benin City, Nigeria*
[e] *Division of Parasitology, National Institute of Cholera and Enteric Diseases, Kolkata 700010, West Bengal, India*
[f] *Enteric Viruses Research Group, Institute of Primate Research, Karen, Nairobi 00502, Kenya*
[g] *College of Animal Science and Veterinary Medicine, Jilin University, 5333 Xi'an Road, Changchun 130062, China*

## ARTICLE INFO

## ABSTRACT

The population structure of *Enterocytozoon bieneusi* was examined by multilocus sequence typing (MLST) of 64 specimens from AIDS patients in Peru, Nigeria, and India and five specimens from captive baboons in Kenya using a combination of the ribosomal internal transcribed spacer (ITS) and four microsatellite and minisatellite markers. Parasites in different geographic locations (Peru, India, and Nigeria) all had strong and significant linkage disequilibrium (LD) and only limited recombination, indicative of a clonal population structure in *E. bieneusi* from each location. When isolates of various geographical areas were treated as a single population, phylogenetic analysis and substructural analysis using STRUCTURE found no evidence for the existence of geographically segregated sub-populations. Nevertheless, both analyses revealed the presence of two major genetically isolated groups of *E. bieneusi*: one (sub-population 1) contained all isolates of the anthroponotic ITS genotype A, whereas the other (sub-population 2) harbored isolates of multiple ITS genotypes with zoonotic potential. This was also supported by $F_{ST}$ analysis. The measurement of LD and recombination rates indicated that sub-population 2 had a clonal population structure, whereas sub-population 1 had an epidemic population structure. The data confirmed the existence of genetic sub-populations in *E. bieneusi* that may be transmitted differently in humans.

Published by Elsevier B.V.

## 1. Introduction

Microsporidia are ubiquitous intracellular pathogens in invertebrate and vertebrate hosts and have been recently reclassified as fungi (Mathis et al., 2005). Of the >14 microsporidian species known to infect humans, *Enterocytozoon bieneusi* is the most common one, infecting many mammals and birds. It is a major cause of opportunistic infections in immunocompromised persons (Akiyoshi et al., 2009; Sokolova et al., 2011). Although *E. bieneusi* is commonly detected in animals and water, little is known of their transmission routes in humans (Anane and Attouchi, 2010; Ghosh and Weiss, 2009; Santin and Fayer, 2011). Because of the lack of

effective therapy against *E. bieneusi*, the elucidation of its epidemiology is a key part of control strategies against microsporidiosis (Anane and Attouchi, 2010).

The ribosomal internal transcribed spacer (ITS) is an effective genetic marker for genotyping *E. bieneusi* and characterizing its transmission routes in humans and animals (Santin and Fayer, 2009). Thus far, over 100 *E. bieneusi* genotypes have been found in multiple hosts, of which many are host-adapted genotypes (Groups II–V) associated with specific animals and probably have no significant public health importance. In contrast, a large group of closely related *E. bieneusi* genotypes (Group I) are frequently found in both humans and animals (Thellier and Breton, 2008). Of the prevalent *E. bieneusi* genotypes (A, D, and IV) in Group I, genotype A was previously identified in a baboon in Kenya, children in Thailand, Portugal, and Peru, villagers in Cameroon, and AIDS patients in Germany, Portugal, Switzerland, Netherlands, Niger, Gabon, and Peru (Li et al., 2011; Lobo et al., 2012; Santin and Fayer, 2011). Genotype D was previously detected in AIDS patients in England, Netherlands, Russia, Portugal, Nigeria, Niger,

Gabon, Cameroon, Malawi, Vietnam, Thailand, Brazil, and Peru, renal transplant recipients in Spain, mice in Germany and Czech Republic, dogs in Portugal, baboons in Kenya, horses in Colombia, falcons in Abu Dhabi, cattle in South Korea and the United States, pigs in Japan, Czech Republic, and the United States, and various wildlife (including foxes, beavers, muskrats, and raccoons) in the United States (Akinbo et al., 2012; Ayinmode et al., 2011; Feng et al., 2011; Galvan et al., 2011; Leelayoova et al., 2006; Li et al., 2011; Lobo et al., 2012; Sak et al., 2011; Saksirisampant et al., 2009; Santin and Fayer, 2011; Sokolova et al., 2011). Genotype IV originally isolated from patients in France was also found in AIDS patients in Gabon, Cameroon, Niger, Nigeria, Malawi, Peru, England, Portugal, and Netherlands, immunocompetent humans in Cameroon, Uganda, Malawi, Nigeria, Portugal, and Netherlands, cats in Germany, Portugal, Japan, and Colombia, dogs in Colombia, and cattle in Portugal, South Korea, and the United States (Akinbo et al., 2012; Ayinmode et al., 2011; Liguory et al., 1998; Lobo et al., 2012; Santin and Fayer, 2011; Sarfati et al., 2006). Because of the uncertainty about the existence of a sexual phase in the *E. bieneusi* lifecycle, the single-marker typing method may have limitations in identifying genotypes that may have different biologic characteristics (Widmer and Akiyoshi, 2010). The single locus typing approach has also failed to address the question concerning whether there is any geographic isolation in *E. bieneusi* genotypes.

Over the years, the use of microsatellite and minisatellite genetic markers in population genetic studies has improved our understanding of the transmission of protozoan parasites (Anderson et al., 2000; Gatei et al., 2007; MacLeod et al., 2000; Morrison et al., 2008; Oura et al., 2005). It has been shown that some parasitic protozoa with a sexual phase in the life cycle can have a clonal population structure, such as *Toxoplasma gondii* and *Cryptosporidium hominis* (Gatei et al., 2007; Howe and Sibley, 1995). However, the population structures of parasites (e.g. *Plasmodium falciparum*, *Trypanosoma brucei*, and *Cryptosporidium parvum*) can vary depending on diverse factors such as host specificity, host movements, geographical regions, and transmission intensity (Mallon et al., 2003; Morrison et al., 2008; Tanriverdi et al., 2006; Tibayrenc and Ayala, 2002). The determination of population genetic structure of microorganisms is vital to the understanding of their transmission patterns (Tibayrenc and Ayala, 2002).

Recently, a multilocus sequencing typing (MLST) tool using four microsatellite and minisatellite markers was developed (Feng et al., 2011), and together with ITS sequencing, it was successfully used to investigate the population structure of 72 *E. bieneusi* specimens from AIDS patients in Peru (Li et al., 2012). It identified two sub-populations of *E. bieneusi* that might differ from each other in genetic structures (clonal and epidemic) and transmission patterns (Li et al., 2012). Nevertheless, these observations on the population genetics of *E. bieneusi* in Peru need to be substantiated in studies done in other geographic locations. In this study, we compared the genetic structure of *E. bieneusi* in AIDS patients in Peru, India, and Nigeria, and determined whether there is any geographic segregation in three common *E. bieneusi* ITS genotypes A, D, and IV.

## 2. Materials and methods

### 2.1. Parasite isolates

The *E. bieneusi* specimens used in the study include 17 from HIV-positive adults in Nigeria, 16 from HIV-positive children in India, and five from captive olive baboons (*Papio anubis*) in Kenya (Table 1). Only one specimen per patient or animal was used in the study. The 17 isolates from Nigeria were genotyped previously as genotypes A (six patients), IV (six patients), D (two patients), Nig2 (one patient), Nig3 (one patient), and Nig5 (one patient)

(Akinbo et al., 2012), the 16 isolates from India as genotypes A (three patients), D (five patients), PigEBITS7 (six patients), and IH (two patients), whereas the five baboon isolates from Kenya as genotype D by ITS sequence analysis (Li et al., 2011) (Table 1). For comparative purposes, MLST data from 26 *E. bieneusi* specimens in HIV-positive adults in Peru from a previous study were included in the population genetic analysis, belonging to nine ITS genotypes: A (five patients), IV (four patients), D (three patients), Peru7 (three patients), Peru11 (three patients), WL11 (three patients), Peru10 (three patients), Peru8 (one patient), and EbpC (one patient) (Table 1) (Li et al., 2012).

### 2.2. MLST PCR and DNA sequence analysis for genetic markers

The five markers used in this study include the ITS, one minisatellite (MS4), and three microsatellites (MS1, MS3, and MS7) as described (Li et al., 2012). Primers and PCR conditions were described previously (Feng et al., 2011; Sulaiman et al., 2003). The secondary PCR products of the anticipated size were purified and sequenced directly with the secondary PCR primers (Li et al., 2012). Sequencing was performed on an ABI 3130 genetic analyzer (Applied Biosystems). The nucleotide sequences obtained were edited and aligned with the sequences of each target obtained in Peru study (Li et al., 2012).

### 2.3. Genetic analysis

The genotypes at each of the five markers and the multilocus genotypes (MLGs) were determined using software DnaSP version 5.10.01 (http://www.ub.edu/dnasp/) based on a combination of sequence length and nucleotide polymorphism (Rozas et al., 2003). We analyzed the population genetics of *E. bieneusi* using data based on multilocus sequences (concatenated from nucleotide sequences of the five markers) and allelic profiles as described (Li et al., 2012). The population structure of *E. bieneusi* was assessed by measuring the intragenic and intergenic linkage disequilibrium (LD), the standardized index of association ($I^S_A$), neutrality, and recombination events (Rms). Recombination rate was estimated using the tests GENECONV, MaxChi, and SiScan implemented in Recombination Detecting Program version 3 (RDP 3–http://darwin.uvigo.es/rdp/rdp.html) as described (Li et al., 2012). The software Arlequin version 3.5.1.2 (http://cmpg.unibe.ch/software/arlequin35/) was used to calculate Wright's fixation index ($F_{ST}$) of isolates from different geographic locations (Excoffier and Lischer, 2010).

### 2.4. Phylogenetic analysis

Software MrBayes version 3.2.1 was used to assess the genetic relationship of multilocus sequences of 64 specimens from four geographic locations (http://mrbayes.sourceforge.net/) (Ronquist and Deans, 2010). A basic Bayesian Markov chain Monte Carlo (MCMC) analysis of phylogeny was conducted using the default parameters settings as follows. Initially, the evolutionary model was set for the general time reversible (GTR) substitution model with gamma-distributed rate variation across sites and a proportion of invariable sites; subsequently, we ran the analysis with a length of 100,000 and sample frequency of 100 and added generations until the average standard deviation of the split frequencies fell below 0.01; this was followed by a summary of the parameter values (mean, mode, and 95% credibility interval) using a burn-in of 250; the final step was to summarize the trees using a burn-in of 250 to generate a cladogram with the posterior probabilities for each split and a phylogram with mean branch lengths. The generated dendrograms were visualized and edited using the software Fig Tree version 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).